

Information Gain versus State Disturbance for a Single Qubit

Konrad Banaszek

*Institute for Quantum Computing, University of Waterloo
Waterloo, ON, N2L 3G1, Canada*

and

*Institute of Physics, Nicolaus Copernicus University
ul. Grudzińska 5, PL-87-100 Toruń, Poland*

e-mail: kbanasz@phys.uni.torun.pl

(Received 2000)

Abstract. The trade-off between the information gain and the state disturbance is derived for quantum operations on a single qubit prepared in a uniformly distributed pure state. The derivation is valid for a class of measures quantifying the state disturbance and the information gain which satisfy certain invariance conditions. This class includes in particular the Shannon entropy versus the operation fidelity. The central role in the derivation is played by efficient quantum operations, which leave the system in a pure output state for any measurement outcome. It is pointed out that the optimality of efficient quantum operations among those inducing a given operator-valued measure is related to Davies' characterization of convex invariant functions on hermitian operators.

1. Introduction

Disturbance caused by an attempt to gain classical information about the quantum state of a physical system is one of the essential features of quantum theory [1, 2, 3, 4, 5, 6, 7, 8]. The balance between the information gain and the state disturbance can be quantified in many different ways, depending on a specific physical scenario. Analytical derivations of the exact form of such trade-offs are usually a non-trivial matter, as the optimization needs to be carried out over all possible quantum operations that can be applied to the system. Nevertheless, in certain instances, for example the case of estimation and operation fidelities for a single d -level system in a uniformly distributed pure state, optimization by analytical means has been shown to be possible [5]. The derivation of this last result was enabled essentially by the fact that both the operation and the estimation fidelities are quadratic in the operators that constitute the Stinespring decomposition of a quantum operation.

The situation becomes considerably more involved when the parameters quantifying the state disturbance or the information gain have a more complicated

dependence on the quantum operation. A good example is the gain of information expressed in terms of the Shannon entropy. This quantity arises naturally in the framework of Bayesian inference, when we try to quantify how well the *a posteriori* probability distribution “pins down” the measured state [9, 10, 11]. In this paper, we present a derivation of the trade-off between the information gain and state disturbance, when the physical system under consideration consists of a single qubit prepared in a pure state, and both the information gain and the state disturbance are quantified using parameters satisfying certain invariance criteria. As a concrete example, we will consider the trade-off between the Shannon entropy as a measure of the information gain combined with the state disturbance quantified in terms of the average operation fidelity.

An intermediate step in this derivation requires demonstrating that it is sufficient to consider the so-called efficient [6] (also known as Lüders-type [4] or ideal [12]) operations, for which obtaining a specific measurement outcome leaves the system in a pure state. Physically speaking, efficient quantum operations should introduce minimum disturbance while inducing a fixed generalized measurement, or equivalently, maximize the amount of classical information for a given Stinespring decomposition of a quantum operation. While the latter property is straightforward to prove for the Shannon information gain or the average estimation fidelity, the former property in the case of the operation fidelity has been conjectured a while ago by Barnum [4], and proven by him only recently [8] using the convexity of certain operator maps involving tensor products [13]. We show here that in the case of uniform distributions generating unitarily invariant measures of the state disturbance, the optimality of efficient operations can be analyzed following an alternative route, based on results obtained in the field of convex analysis on operators [14, 15, 16, 17]. This connection is valid for arbitrary finite-dimensional systems, and it is an interesting question whether any of these results can be carried over to Hilbert spaces of infinite dimension.

This paper is structured as follows. First we discuss the convexity properties of measures characterizing information gain and state disturbance in Sect. 2 and 3 using the examples of the Shannon information gain and the operation fidelity. In Sect. 3 we also link the optimality of efficient quantum operations to the convexity of unitarily invariant functions defined on hermitian operators. In Sect. 4 we derive the trade-off between the information gain and the state disturbance for a single qubit. Sect. 5 gives two other measures of information gain and state disturbance which fall under the general formalism used in the preceding section. Finally, Sect. 6 concludes the paper.

Before we pass on to detailed discussion, let us first summarize briefly the elementary facts needed further in this paper. The subject of our interest is a quantum operation, i.e. a trace-preserving completely positive map acting on density

operators $\hat{\varrho}$. The map has the Stinespring decomposition of the form

$$\hat{\varrho} \mapsto \sum_{r\mu} \hat{A}_{r\mu} \hat{\varrho} \hat{A}_{r\mu}^\dagger. \quad (1)$$

with the family of the operators $\{\hat{A}_{r\mu}\}$ satisfying the condition

$$\sum_{r\mu} \hat{A}_{r\mu}^\dagger \hat{A}_{r\mu} = \hat{\mathbb{I}} \quad (2)$$

which guarantees the preservation of the trace of the density operator. We assume that the index r describes the classical outcome of the measurement, while the summation over the index μ is responsible for additional averaging due to the imperfections of the measuring apparatus, resulting in a decrease of classically available information. If we are interested only in the classical outcome of the measurement, then it is sufficient to consider a positive operator-valued measure $\{\hat{M}_r\}$ induced by the quantum operation $\{\hat{A}_{r\mu}\}$ according to

$$\hat{M}_r = \sum_{\mu} \hat{A}_{r\mu}^\dagger \hat{A}_{r\mu}. \quad (3)$$

The probability of obtaining a specific result r is given by $\text{Tr}(\hat{M}_r \hat{\varrho})$, and the outcome r is associated with a conditional transformation of the density matrix given by

$$\hat{\varrho} \mapsto \frac{1}{\text{Tr}(\hat{M}_r \hat{\varrho})} \sum_{\mu} \hat{A}_{r\mu} \hat{\varrho} \hat{A}_{r\mu}^\dagger. \quad (4)$$

The gain of classical information depends only on the operator-valued measure $\{\hat{M}_r\}$, whereas the state disturbance is affected by the specific form of the quantum operation. Of course, there are many quantum operations corresponding to a given operator-valued measure $\{\hat{M}_r\}$. Given an operation $\{\hat{A}_{r\mu}\}$ we can always augment the associated information gain by assuming that both the indices r and μ are available classically. Conversely, we can ask which quantum operation inducing the generalized measurement $\{\hat{M}_r\}$ minimizes the state disturbance. A good candidate for this operation is $\{\sqrt{\hat{M}_r}\}$, as it retains the purity of the input state for a given measurement outcome, and also preserves the relative phases of the input state in the basis diagonalizing \hat{M}_r . As it will be discussed in Sect. 3, this intuition turns out to be correct when the measure of the state disturbance can be represented as a sum of contributions given by the values of a certain function on the operators $\hat{A}_{r\mu}^\dagger \hat{A}_{r\mu}$, when this function is unitarily invariant and concave on the subspace of diagonal operators.

2. Shannon information gain

As an example of a measure of classical information gained from a quantum operation, we shall consider here the decrease of the Shannon entropy between the *a priori* and *a posteriori* probability distributions defined on the manifold of pure states. Let us denote by $d\psi$ the normalized invariant integration measure on the manifold of pure states. This measure is induced by the Haar measure associated with the Lie group of unitary transformations on the corresponding Hilbert space. We will assume that the initial *a priori* probability distribution is given with respect to this measure by a function $p(\psi)$. Therefore the initial entropy is given by the expression

$$H[p(\psi)] = - \int d\psi p(\psi) \log_2 p(\psi). \quad (5)$$

Obtaining a specific outcome r of the measurement yields a conditional *a posteriori* probability distribution $p(\psi|r)$ calculated according to the Bayes' rule as

$$p(\psi|r) = \frac{p(r|\psi)p(\psi)}{p(r)} \quad (6)$$

where $p(r|\psi) = \langle \psi | \hat{M}_r | \psi \rangle$ is the standard quantum mechanical probability of obtaining the outcome r for an input state $|\psi\rangle$, and $p(r)$ is the average:

$$p(r) = \int d\psi p(r|\psi)p(\psi) \quad (7)$$

The Shannon entropy of the conditional *a posteriori* distribution $p(\psi|r)$ is given by the expression $H[p(\psi|r)]$. As we are interested in the average decrease of the entropy after the quantum operation is carried out, we need to take the difference $H[p(\psi)] - H[p(\psi|r)]$ and average it over all possible outcomes of the experiment with the probability distribution $p(r)$. Thus the final expression for the average gain of the Shannon information H from a measurement described by a positive operator-valued measure $\{\hat{M}_r\}$ is given by:

$$\begin{aligned} H &= \sum_r p(r) \{H[p(\psi)] - H[p(\psi|r)]\} \\ &= \sum_r \int d\psi p(r|\psi)p(\psi) \log_2 \frac{p(r|\psi)}{p(r)} \\ &= \sum_r \int d\psi p(\psi) \langle \psi | \hat{M}_r | \psi \rangle \log_2 \frac{\langle \psi | \hat{M}_r | \psi \rangle}{\int d\psi p(\psi) \langle \psi | \hat{M}_r | \psi \rangle}. \end{aligned} \quad (8)$$

In the following, it will be convenient to represent the average information gain as a sum

$$H = \sum_r \mathcal{H}(\hat{M}_r), \quad (9)$$

where $\mathcal{H}(\hat{M})$ is a function defined for semipositive operators \hat{M} according to:

$$\mathcal{H}(\hat{M}) = \int d\psi p(\psi) \langle \psi | \hat{M} | \psi \rangle \log_2 \frac{\langle \psi | \hat{M} | \psi \rangle}{\int d\psi p(\psi) \langle \psi | \hat{M} | \psi \rangle} \quad (10)$$

Let us note that $\mathcal{H}(\hat{M})$ is positively homogeneous of degree one.

Physically, we anticipate that combining two elements \hat{M}_1 and \hat{M}_2 of the operator measure into one operator $\hat{M}_1 + \hat{M}_2$ should result in the loss of information. Such a loss of information should correspond to the inequality

$$\mathcal{H}(\hat{M}_1) + \mathcal{H}(\hat{M}_2) \geq \mathcal{H}(\hat{M}_1 + \hat{M}_2), \quad (11)$$

which is equivalent to the convexity of $\mathcal{H}(\hat{M})$ due to the fact that $\mathcal{H}(\hat{M})$ is positively homogeneous of degree one. The above inequality is a general property of mutual information [18], and it follows elementarily from the convexity of the function $x \log_2 x$ on the interval $[0, 1]$:

$$tx_1 \log_2 x_1 + (1-t)x_2 \log_2 x_2 \geq [tx_1 + (1-t)x_2] \log_2 [tx_1 + (1-t)x_2] \quad (12)$$

Inserting

$$t = \frac{p(r_1)}{p(r_1) + p(r_2)}, \quad x_1 = \frac{p(r_1|\psi)}{p(r_1)}, \quad x_2 = \frac{p(r_2|\psi)}{p(r_2)} \quad (13)$$

yields the convexity of the integrand in (10), and integrating it with $\int d\psi p(\psi)$ proves the convexity of \mathcal{H} . By mathematical induction, we therefore obtain that for a quantum operation $\{\hat{A}_{r\mu}\}$ the information gain is maximized by assuming that both the indices $r\mu$ are available classically, and by constructing a “finer” generalized measurement from the quantum operation as $\hat{M}_{r\mu} = \hat{A}_{r\mu}^\dagger \hat{A}_{r\mu}$. An analogous statement holds also for any other measure of information gain which is decomposable to the form (9) with a positively homogeneous convex function \mathcal{H} .

3. Operation fidelity

If the initial state $|\psi\rangle$ of the system is chosen according to a probability distribution $p(\psi)$, then the operation fidelity F for a quantum operation $\{\hat{A}_{r\mu}\}$, defined as the average projection of the final density matrix onto the initial pure state, is given by the expression:

$$F = \sum_{r\mu} \int d\psi p(\psi) |\langle \psi | \hat{A}_{r\mu} | \psi \rangle|^2. \quad (14)$$

For a uniform distribution of input states with $p(\psi) = 1$ the integral $\int d\psi$ can be carried out analytically in a d -dimensional Hilbert space [4, 19], yielding the following expression for the mean operation fidelity [4, 5]:

$$F = \frac{1}{d(d+1)} \sum_{r\mu} [\text{Tr}(\hat{A}_{r\mu}^\dagger \hat{A}_{r\mu}) + |\text{Tr} \hat{A}_{r\mu}|^2]. \quad (15)$$

By applying the singular value decomposition to operators $\hat{A}_{r\mu}$, it can be shown by elementary means that $\text{Tr} \hat{A}_{r\mu} \leq \text{Tr} \sqrt{\hat{A}_{r\mu}^\dagger \hat{A}_{r\mu}}$, and consequently the operation fidelity can only increase when replacing $\hat{A}_{r\mu}$ with $\sqrt{\hat{A}_{r\mu}^\dagger \hat{A}_{r\mu}}$, while the generalized measurement induced by the operation (and consequently any measure of the information gain) remain the same. This implies that it is sufficient to restrict our considerations to quantum operations composed of semipositive definite hermitian operators $\hat{A}_{r\mu}$. In this case, we can represent the operation fidelity as a sum over $r\mu$ of a certain function \mathcal{F} of the products $\hat{A}_{r\mu}^\dagger \hat{A}_{r\mu}$:

$$F = \sum_{r\mu} \mathcal{F}(\hat{A}_{r\mu}^\dagger \hat{A}_{r\mu}). \quad (16)$$

The function $\mathcal{F}(\hat{M})$, given explicitly by:

$$\mathcal{F}(\hat{M}) = \frac{1}{d(d+1)} (\text{Tr} \hat{M} + |\text{Tr} \sqrt{\hat{M}}|^2) \quad (17)$$

is well defined for all semipositive hermitian operators \hat{M} , which form a convex set. Let us note that similarly to the function $\mathcal{H}(\hat{M})$ introduced in the preceding section it is also positively homogeneous of degree one.

We would like now to find a quantum operation which maximizes the mean operation fidelity under the constraint of inducing a specified operator-valued measure according to (3). Recalling the introductory discussion, we expect that the optimal operation is given by square roots of the operator measure $\{\sqrt{\hat{M}_r}\}$. In order to obtain this result, it is sufficient to show that for any semipositive definite operators \hat{M}_1 and \hat{M}_2 the function $\mathcal{F}(\hat{M})$ satisfies the following inequality:

$$\mathcal{F}(\hat{M}_1) + \mathcal{F}(\hat{M}_2) \leq \mathcal{F}(\hat{M}_1 + \hat{M}_2), \quad (18)$$

which again is equivalent to its concavity, as $\mathcal{F}(\hat{M})$ is positively homogeneous. Assuming that (18) holds, we would immediately obtain that for any two elements $\hat{A}_{r\mu_1}$ and $\hat{A}_{r\mu_2}$ of a quantum operation we have: $\mathcal{F}(\hat{A}_{r\mu_1}^\dagger \hat{A}_{r\mu_1}) + \mathcal{F}(\hat{A}_{r\mu_2}^\dagger \hat{A}_{r\mu_2}) \leq \mathcal{F}(\hat{A}_{r\mu_1}^\dagger \hat{A}_{r\mu_1} + \hat{A}_{r\mu_2}^\dagger \hat{A}_{r\mu_2})$, and by induction:

$$\sum_{\mu} \mathcal{F}(\hat{A}_{r\mu}^\dagger \hat{A}_{r\mu}) \leq \mathcal{F} \left(\sum_{\mu} \hat{A}_{r\mu}^\dagger \hat{A}_{r\mu} \right) = \mathcal{F}(\hat{M}_r), \quad (19)$$

which would prove our thesis.

In order to demonstrate (18), let us first note that \mathcal{F} is invariant with respect to unitary transformations of its argument, i.e. $\mathcal{F}(\hat{U}^\dagger \hat{M} \hat{U}) = \mathcal{F}(\hat{M})$ for any semipositive \hat{M} and unitary \hat{U} . Therefore $\mathcal{F}(\hat{M})$ can depend only on the eigenvalues

of \hat{M} . Let us consider a function ϕ defined on d -dimensional vectors $\mathbf{u} \in \mathbb{R}_+^d$ with nonnegative coordinates, given by the value of \mathcal{F} on a diagonal operator $\text{diag}(\mathbf{u})$:

$$\phi(\mathbf{u}) = \mathcal{F}(\text{diag}(\mathbf{u})). \quad (20)$$

In the case of \mathcal{F} given by (15), the function ϕ reads:

$$\phi(\mathbf{u}) = \frac{1}{d(d+1)} \left(\sum_{i=0}^{d-1} u_i + \left(\sum_{i=0}^{d-1} \sqrt{u_i} \right)^2 \right) \quad (21)$$

and it is symmetric, i.e. invariant with respect to the permutations of the coordinates u_0, u_1, \dots, u_{d-1} of the vector \mathbf{u} . It is easy to verify that for any pair \mathbf{u}, \mathbf{v} the following inequality holds:

$$\phi(\mathbf{u}) + \phi(\mathbf{v}) \leq \phi(\mathbf{u} + \mathbf{v}), \quad (22)$$

which analogously as before means that ϕ is concave due to its positive homogeneity. Indeed, by writing explicitly the left and the right hand sides of the above inequality we have:

$$\sum_{i,j=0}^{d-1} \sqrt{u_i u_j} + \sqrt{v_i v_j} \leq \sum_{i,j=0}^{d-1} \sqrt{(u_i + v_i)(u_j + v_j)} \quad (23)$$

and it is straightforward to see that the inequality holds separately for every term with fixed i and j , which proves (22). The inequality (22) immediately implies the inequality (18) for pairs of operators \hat{M}_1 and \hat{M}_2 which commute and therefore can be diagonalized simultaneously in the same orthonormal basis.

The definition (20) introduces a correspondence between a unitarily invariant function \mathcal{F} and a function ϕ which is symmetric, i.e. invariant with respect to the permutations of the coordinates of its argument. The crucial step of the reasoning is the application of a theorem proven first by Davies [14] and by Friedland [15] stating that in this setting the convexity (concavity) of \mathcal{F} is equivalent to the convexity (concavity) of ϕ on their respective domains. Hence (22) implies the inequality (18) also for non-commuting pairs of operators \hat{M}_1 and \hat{M}_2 . This immediately proves that for a given generalized measurement $\{\hat{M}\}$ the quantum operation minimizing the state disturbance quantified with the mean operation fidelity is given by $\{\sqrt{\hat{M}}\}$ in the case of a uniform distribution on pure input states.

4. Trade-off for a single qubit

The results described in the preceding sections allow us to restrict the search for the information gain versus state disturbance trade-off to efficient quantum

operations generated from positive operator-valued measures $\{\hat{M}_r\}$ by taking the square roots of its elements $\{\sqrt{\hat{M}_r}\}$. Furthermore, we will consider \mathcal{F} and \mathcal{H} which are respectively concave and convex unitarily invariant functions with the property of positive homogeneity. For a single qubit, these assumptions severely restrict the number of parameters characterizing the elements \hat{M}_r of the operator-valued measure which play a non-trivial role in the derivation. In fact, as we will see in a moment, the only relevant parameter is the ratio of the eigenvalues of \hat{M}_r . Let us start by using the positive homogeneity of \mathcal{F} and \mathcal{H} to write:

$$\begin{aligned}\mathcal{F}(\hat{M}_r) &= \frac{1}{2} \text{Tr}(\hat{M}_r) \mathcal{F}\left(\frac{2\hat{M}_r}{\text{Tr}(\hat{M}_r)}\right) \\ \mathcal{H}(\hat{M}_r) &= \frac{1}{2} \text{Tr}(\hat{M}_r) \mathcal{H}\left(\frac{2\hat{M}_r}{\text{Tr}(\hat{M}_r)}\right)\end{aligned}\tag{24}$$

and denote $\xi_r = \text{Tr}(\hat{M}_r)/2$. Of course $\xi_r \geq 0$, and taking the trace of $\sum_r \hat{M}_r = \hat{\mathbb{1}}$ yields a summation condition on $\{\xi_r\}$:

$$\sum_r \xi_r = 1.\tag{25}$$

Next, let us denote the eigenvalues of the renormalized operators $2\hat{M}_r/\text{Tr}(\hat{M}_r)$ as:

$$\frac{2}{\text{Tr}(\hat{M}_r)} \hat{M}_r = \begin{pmatrix} 1+x_r & 0 \\ 0 & 1-x_r \end{pmatrix}.\tag{26}$$

The positivity of the eigenvalues requires that $-1 \leq x_r \leq 1$, and if we further assume with no loss of generality that the eigenvalues are arranged in a non-increasing order, we can restrict our interest to the range $0 \leq x_r \leq 1$. In the new variables we can now write:

$$\begin{aligned}\mathcal{F}(\hat{M}_r) &= \xi_r f(x_r) \\ \mathcal{H}(\hat{M}_r) &= \xi_r h(x_r)\end{aligned}\tag{27}$$

where the functions $f, h : [0, 1] \rightarrow \mathbb{R}$ are defined as:

$$f(x) = \mathcal{F}(\text{diag}(1+x, 1-x))\tag{28}$$

$$h(x) = \mathcal{H}(\text{diag}(1+x, 1-x))\tag{29}$$

It is straightforward to show that the concavity of \mathcal{F} and the convexity of \mathcal{H} together with their positive homogeneity imply that the functions f and h are respectively concave and convex. Furthermore, the converse is also true.

Physically, we expect that f is strictly decreasing, whereas h is strictly increasing on their domain $[0, 1]$. The reason for this is that the parameter x characterizes

the imbalance between the eigenvalues of \hat{M}_r for a fixed trace $\text{Tr}(\hat{M}_r)$, and balanced pairs of eigenvalues do not disturb the state, whereas unbalanced pairs of eigenvalues are responsible for the information gain. If we now assume that the function f is strictly monotonic, then the information gain H can be expressed as:

$$H = \sum_r \xi_r h(x_r) = \sum_r \xi_r h(f^{-1}(f(x_r))) = \sum_r \xi_r (h \circ f^{-1})(f(x_r)) \quad (30)$$

Let us now recall the notion of a concave envelope [20] of a real function $\chi : C \rightarrow \mathbb{R}$ defined on a convex set C . The concave envelope, which we will denote by $\bar{\chi} : C \rightarrow \mathbb{R}$ is defined as a concave function such that $\bar{\chi}(x) \geq \chi(x)$ for all $x \in C$, and also that for any other concave function $\tilde{\chi} : C \rightarrow \mathbb{R}$ satisfying $\tilde{\chi}(x) \geq \chi(x)$ everywhere we have also $\bar{\chi}(x) \geq \tilde{\chi}(x)$ everywhere on C . In the one-dimensional case the concave envelope is given explicitly by the expression

$$\bar{\chi}(x) = \sup_{t, x_1, x_2} \{t\chi(x_1) + (1-t)\chi(x_2) \mid 0 \leq t \leq 1, tx_1 + (1-t)x_2 = x\}. \quad (31)$$

In our derivation, we will take $\chi = h \circ f^{-1}$, and x_1, x_2 in the above formula are any two points from the image of f . The concave envelope $\bar{\chi} = \bar{h} \circ f^{-1}$ can be used to estimate H in (30) from above, and furthermore the concavity of $\bar{\chi}$ allows us to apply Jensen's inequality. Combining these two steps yields:

$$H \leq \sum_r \xi_r \bar{\chi}(f(x_r)) \leq \bar{\chi}\left(\sum_r \xi_r f(x_r)\right) = \bar{\chi}(F). \quad (32)$$

Thus finally

$$H \leq \bar{h} \circ f^{-1}(F), \quad (33)$$

where $\bar{h} \circ f^{-1}$ is defined according to (31). Let us note that in general the concavity of f and the convexity of h along with their monotonicity do not guarantee automatically that the composition $h \circ f^{-1}$ is concave itself, hence the need to take its concave envelope. This can be seen most easily when both f and h are doubly differentiable on $[0, 1]$. Then we have:

$$(h \circ f^{-1})'' = \frac{h'' f' - h' f''}{(f')^2} \quad (34)$$

and the conditions $h' > 0$, $h'' \geq 0$, $f' < 0$, and $f'' \leq 0$ do not imply a well-defined sign of the right-hand side of the above formula.

The inequality (33) can be easily saturated. If we take a two-element quantum operation of the form

$$\{\hat{A}_1(x) = \text{diag}(\sqrt{(1+x)/2}, \sqrt{(1-x)/2}), \hat{A}_2(x) = \text{diag}(\sqrt{(1-x)/2}, \sqrt{(1+x)/2})\}$$

with $0 \leq x \leq 1$, we can generate any point of the graph of the function $h \circ f^{-1}$. Furthermore, by taking suitable combinations of two quantum operations that have the above form

$$\{\sqrt{t}\hat{A}_1(x_1), \sqrt{t}\hat{A}_2(x_1), \sqrt{1-t}\hat{A}_1(x_2), \sqrt{1-t}\hat{A}_2(x_2)\}$$

with $0 \leq t \leq 1$ we can generate an arbitrary point of the graph of the function $h \circ f^{-1}$, which follows from (31). Let us also note that if the composition $h \circ f^{-1}$ is a strictly concave function, then the equality sign in (33) is reached if and only if the parameters x_r are equal. This means that for a strictly concave $h \circ f^{-1}$ all the elements of an operator valued measure saturating the trade-off need to have the same ratio of their eigenvalues.

Let us now specialize the above result to the trade-off expressed in terms of the operation fidelity versus the Shannon information gain. In the case of the average operation fidelity, the function f defined in (28) has the following explicit form:

$$f(x) = \frac{1}{3}(2 + \sqrt{1 - x^2}) \quad (35)$$

Furthermore, as sketched in Appendix A, it is straightforward to obtain a closed expression for the Shannon information:

$$h(x) = \frac{1+x^2}{4x} \log_2 \left(\frac{1+x}{1-x} \right) + \frac{1}{2} \log_2(1-x^2) - \frac{1}{2 \ln 2} \quad (36)$$

which is a particular case of the general result obtained by Jones [21]. In the case of f and h given by (35) and (36) the composition $h \circ f^{-1}$ is strictly concave on the whole image of f , which is also shown in Appendix A. Consequently, inverting the function f given in (35) brings the trade-off inequality:

$$H \leq h(\sqrt{1 - (3F - 2)^2}) \quad (37)$$

with h given by (36). As the right-hand side is strictly concave over the domain of F , Jensen's inequality is saturated only for operations for which all x_r are equal. This means that the ratio of the eigenvalues for all the elements of the operator measure $\{\hat{M}_r\}$ needs to be constant across the index r .

5. Other measures

In order to illustrate the generality of the derivation presented in the preceding section, let us discuss two other measures of information gain and state disturbance for which the above reasoning leading to the trade-off for a single qubit holds as well.

As an example of another measure of the information gain satisfying all the above properties let us recall the example of the mean estimation fidelity [22, 23, 24]

for which the trade-off against the operation fidelity has been studied in [5, 7]. Using the notation of the present paper we can write the estimation fidelity for a uniform input distribution in d dimensions as a sum $G = \sum_r \mathcal{G}(\hat{M}_r)$ of terms given by:

$$\mathcal{G}(\hat{M}) = \frac{1}{d(d+1)} [\text{Tr}(\hat{M}) + \|\sqrt{\hat{M}}\|^2] \quad (38)$$

where $\|\cdot\|$ stands for the standard Euclidean operator norm. The above function again is positively homogeneous and convex:

$$\mathcal{G}(\hat{M}_1) + \mathcal{G}(\hat{M}_2) \geq \mathcal{G}(\hat{M}_1 + \hat{M}_2), \quad (39)$$

the latter property following immediately from the lower bound on the second term in (38) for a sum of two semipositive definite hermitian operators \hat{M}_1 and \hat{M}_2 :

$$\begin{aligned} \|\sqrt{\hat{M}_1 + \hat{M}_2}\|^2 &= \sup_{\langle \phi | \phi \rangle = 1} \langle \phi | \hat{M}_1 + \hat{M}_2 | \phi \rangle = \langle \phi_0 | \hat{M}_1 + \hat{M}_2 | \phi_0 \rangle \\ &= \langle \phi_0 | \hat{M}_1 | \phi_0 \rangle + \langle \phi_0 | \hat{M}_2 | \phi_0 \rangle \\ &\leq \sup_{\langle \phi | \phi \rangle = 1} \langle \phi | \hat{M}_1 | \phi \rangle + \sup_{\langle \phi | \phi \rangle = 1} \langle \phi | \hat{M}_2 | \phi \rangle \\ &= \|\sqrt{\hat{M}_1}\|^2 + \|\sqrt{\hat{M}_2}\|^2, \end{aligned} \quad (40)$$

where $|\phi_0\rangle$ is the eigenvector of $\hat{M}_1 + \hat{M}_2$ corresponding to its largest eigenvalue.

Specialized to the qubit system, \mathcal{G} induces the function g :

$$g = \mathcal{G}(\text{diag}(1+x, 1-x)) = \frac{1}{6}(x+3). \quad (41)$$

As an alternative measure of the state disturbance we will consider the absolute value of the scalar product between the input and the output state, and its generalization to density matrices [25, 26, 27]. In this paper we will call it the Bures-Uhlmann fidelity in order to distinguish it from the fidelity discussed in Sect. 3, which is a square of the former [28]. For a distribution on pure input states characterized by a probability distribution $p(\psi)$, the average Bures-Uhlmann fidelity is given by

$$B = \sum_{r\mu} \int d\psi p(\psi) \sqrt{\langle \psi | \hat{A}_{r\mu}^\dagger \hat{A}_{r\mu} | \psi \rangle} |\langle \psi | \hat{A}_{r\mu} | \psi \rangle|. \quad (42)$$

The demonstration that quantum operations composed of semipositive definite hermitian operators optimize the Bures-Uhlmann fidelity turns out to be substantially more complicated than in the case of F . The particular case of a single qubit with

$p(\psi) = 1$ is discussed in Appendix B. If we restrict our attention to quantum operations composed of hermitian and semipositive definite operators $\hat{A}_{r\mu} = \sqrt{\hat{A}_{r\mu}^\dagger \hat{A}_{r\mu}}$, we can write B as a sum:

$$B = \sum_{r\mu} \mathcal{B}(\hat{A}_{r\mu}^\dagger \hat{A}_{r\mu}) \quad (43)$$

where the function \mathcal{B} is defined on hermitian semipositive operators by the equation:

$$\mathcal{B}(\hat{M}) = \int d\psi p(\psi) \sqrt{\langle \psi | \hat{M} | \psi \rangle} \langle \psi | \sqrt{\hat{M}} | \psi \rangle. \quad (44)$$

The function \mathcal{B} is positively homogeneous, and it is also concave for commuting operators \hat{M}_1 and \hat{M}_2 . In order to show this, let us introduce an orthonormal basis $|i\rangle$ in which both the operators are diagonal: $\hat{M}_1 = \sum_{i=0}^{d-1} u_i |i\rangle\langle i|$ and $\hat{M}_2 = \sum_{i=0}^{d-1} v_i |i\rangle\langle i|$, and denote $p_i = |\langle i | \psi \rangle|^2$. By treating $\sqrt{u_i}$, $\sqrt{v_i}$, $\sqrt{\sum_{j=0}^{d-1} p_j u_j}$, and $\sqrt{\sum_{j=0}^{d-1} p_j v_j}$ as four independent nonnegative numbers, it is straightforward to show that

$$\sqrt{u_i} \sqrt{\sum_{j=0}^{d-1} p_j u_j} + \sqrt{v_i} \sqrt{\sum_{j=0}^{d-1} p_j v_j} \leq \sqrt{u_i + v_i} \sqrt{\sum_{j=0}^{d-1} p_j (u_j + v_j)} \quad (45)$$

This inequality, summed over i with the weights p_i yields:

$$\begin{aligned} \langle \psi | \sqrt{\hat{M}_1} | \psi \rangle \sqrt{\langle \psi | \hat{M}_1 | \psi \rangle} + \langle \psi | \sqrt{\hat{M}_2} | \psi \rangle \sqrt{\langle \psi | \hat{M}_2 | \psi \rangle} \\ \leq \langle \psi | \sqrt{\hat{M}_1 + \hat{M}_2} | \psi \rangle \sqrt{\langle \psi | \hat{M}_1 + \hat{M}_2 | \psi \rangle} \end{aligned} \quad (46)$$

which integrated over pure states with $\int d\psi p(\psi)$ implies that for any two commuting operators \hat{M}_1 and \hat{M}_2 we have:

$$\mathcal{B}(\hat{M}_1) + \mathcal{B}(\hat{M}_2) \leq \mathcal{B}(\hat{M}_1 + \hat{M}_2). \quad (47)$$

Furthermore, if $p(\psi) = 1$ then \mathcal{B} is unitarily invariant, and we can again use Davies' result to show that \mathcal{B} is concave for all, not necessarily commuting, pairs of semipositive hermitian operators.

The above result means in particular that for a qubit prepared in a uniformly distributed pure state efficient operations composed of hermitian semipositive elements are optimal from the point of view of the trade-off. In this case it is sufficient to consider a function b of a single real parameter x defined analogously to Sect. 4 as:

$$\begin{aligned} b(x) &= \mathcal{B}(\text{diag}(x+1, x-1)) \\ &= \frac{2}{15x^2} [(1+x^2)\sqrt{1-x^2} + 7x^2 - 1]. \end{aligned} \quad (48)$$

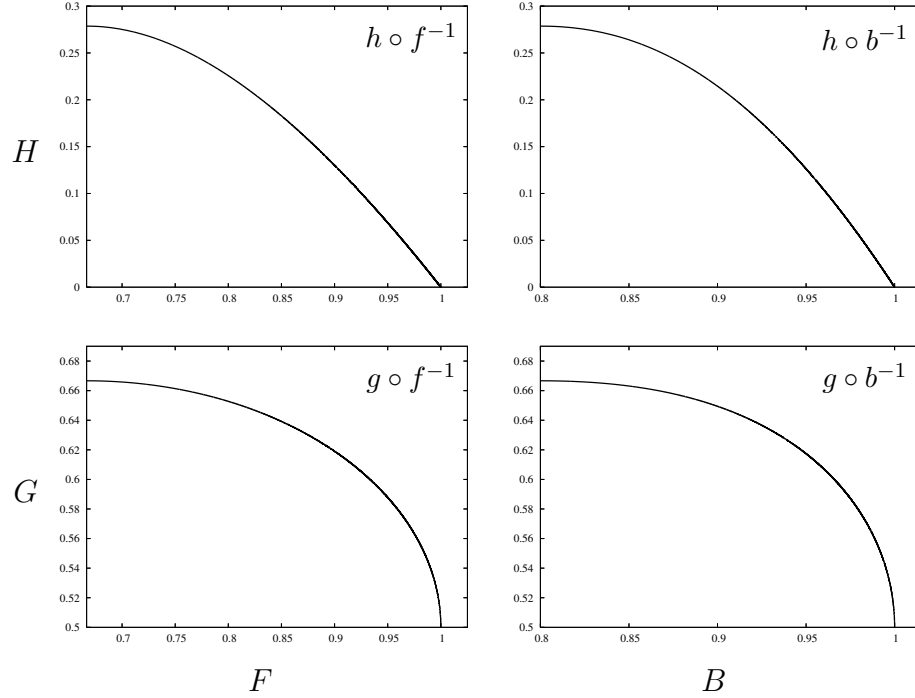


Fig. 1: The quantum mechanical trade-offs between the information gain and the state disturbance with the state disturbance quantified as the average operation fidelity F and the average Bures-Uhlmann fidelity B , and the information gain measured using the change H in Shannon entropy, and the average estimation fidelity G . In all four cases the trade-offs are characterized by concave composite maps specified in the upper right corners of the graphs.

The resulting trade-offs for any combination of F and B with H and G are depicted in Fig. 1. It is seen that in all four cases the composite maps characterizing the trade-offs are concave themselves and therefore equal to their concave envelopes.

6. Conclusions

We have discussed the convexity properties of measures quantifying the information gain and the state disturbance in quantum operations, using the examples of the average operation fidelity and the Shannon entropy. Such convexity properties can be expected from all physically motivated measures, and in the case of uniform *a priori* distribution which implies invariance with respect to unitary

transformations, we can use the theory of convex invariant functions to analyze the convexity. In the case of quantum operations on a single qubit, the trade-offs can be described using a general inequality, resulting from the existence of a single relevant parameter characterizing the elements of the positive operator-valued measure through the ratio of their eigenvalues.

Acknowledgements

I have benefited from exchanging ideas with H. Barnum, I. Devetak, and C. A. Fuchs. I am grateful to R. Cleve, R. Laflamme, and M. Mosca for discussions and their hospitality during my stay at the Institute of Quantum Computing of the University of Waterloo, whose splendid library resources made me familiar with Refs. [14, 15, 16, 17]. This research was supported by MNI project no. 1 P03B 011 29.

Appendix A

In order to evaluate the function h defined in (29) let us introduce the standard parameterization of the manifold of the single-qubit pure states:

$$|\psi\rangle = \begin{pmatrix} \cos(\theta/2) \\ e^{i\varphi} \sin(\theta/2) \end{pmatrix}, \quad d\psi = \frac{1}{4\pi} \sin \theta \, d\theta \, d\varphi, \quad (49)$$

with $0 \leq \theta \leq \pi$ and $0 \leq \varphi \leq 2\pi$. In this parametrization we have:

$$\begin{aligned} h(x) &= \mathcal{H}(\text{diag}(1+x, 1-x)) \\ &= \frac{1}{2} \int_0^\pi d\theta \sin \theta (1+x \cos \theta) \log_2(1+x \cos \theta) \\ &= \frac{1+x^2}{4x} \log_2 \left(\frac{1+x}{1-x} \right) + \frac{1}{2} \log_2(1-x^2) - \frac{1}{2 \ln 2}. \end{aligned} \quad (50)$$

In order to demonstrate the concavity of $h \circ f^{-1}$ we need to check the concavity of the function $h(\sqrt{1-x^2})$ on the interval $x \in [0, 1]$. The second derivative of $h(\sqrt{1-x^2})$ is given by

$$\frac{d^2}{dx^2} h(\sqrt{1-x^2}) = -\frac{3x^2}{4(1-x^2)^{5/2}} \left[\log_2 \left(\frac{1-\sqrt{1-x^2}}{1+\sqrt{1-x^2}} \right) + \frac{(4x^2+2)\sqrt{1-x^2}}{3x^2 \ln 2} \right] \quad (51)$$

The expression in the square parentheses is equal to 0 for $x = 0$, and then its derivative is equal to $-4(1-x^2)^{3/2}/3x^2 \ln 2$, which is strictly negative for $x \in]0, 1[$. This implies that the expression under consideration takes negative values on that interval. Consequently $d^2 h(\sqrt{1-x^2})/dx^2$ is negative for $x \in]0, 1[$.

Appendix B

Let us consider the contribution to the Bures-Uhlmann fidelity generated by an element of quantum operation of the form $\hat{U}\hat{A}$ where \hat{U} is unitary and $\hat{A} = \text{diag}(\sqrt{1+x}, \sqrt{1-x})$ with $x \in [0, 1]$. It will be convenient to switch to the Bloch vector representation, where the density matrix of a pure state $|\psi\rangle\langle\psi|$ is represented by a three-dimensional real vector \mathbf{r} according to:

$$|\psi\rangle\langle\psi| = \frac{1}{2}(\hat{\mathbb{1}} + \mathbf{r}^T \cdot \hat{\boldsymbol{\sigma}}) \quad (52)$$

where $\hat{\boldsymbol{\sigma}}$ is a vector composed of the three Pauli matrices. If we denote $\mathbf{A} = \text{diag}(\sqrt{1-x^2}, \sqrt{1-x^2}, 1)$ and $\mathbf{a} = (0, 0, x)^T$, then the expectation value $\langle\psi|\hat{A}^\dagger\hat{A}|\psi\rangle$ can be written as:

$$\langle\psi|\hat{A}^\dagger\hat{A}|\psi\rangle = 2(1 + \mathbf{a}^T \cdot \mathbf{r}), \quad (53)$$

where the dot \cdot denotes matrix multiplication, and the conditional transformation $|\psi\rangle\langle\psi| \mapsto \hat{U}\hat{A}|\psi\rangle\langle\psi|\hat{A}^\dagger\hat{U}^\dagger / \langle\psi|\hat{A}^\dagger\hat{A}|\psi\rangle$ is given in the Bloch representation by

$$\mathbf{r} \mapsto \frac{\mathbf{O} \cdot (\mathbf{A} \cdot \mathbf{r} + \mathbf{a})}{1 + \mathbf{a}^T \cdot \mathbf{r}} \quad (54)$$

where \mathbf{O} is the rotation of the Bloch vector corresponding to the unitary transformation \hat{U} .

The integrand in (42) for $p(\psi) = 1$ can now be written as

$$\sqrt{\langle\psi|\hat{A}^\dagger\hat{A}|\psi\rangle} |\langle\psi|\hat{U}\hat{A}|\psi\rangle| = \frac{1}{\sqrt{2}} \sqrt{(1 + \mathbf{a}^T \cdot \mathbf{r})^2 + (1 + \mathbf{a}^T \cdot \mathbf{r}) \mathbf{r}^T \cdot \mathbf{O} \cdot (\mathbf{A} \cdot \mathbf{r} + \mathbf{a})}. \quad (55)$$

Following (49), we will parametrize the Bloch vector \mathbf{r} as

$$\mathbf{r} = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)^T \quad (56)$$

and then the Bures-Uhlmann fidelity for the operator $\hat{U}\hat{A}$ is the expression given in (55) integrated over

$$\int d^2\mathbf{r} = \frac{1}{2} \int_0^\pi \sin \theta d\theta \int_0^{2\pi} \frac{d\varphi}{2\pi}. \quad (57)$$

Let us first consider the second integral over the azimuthal angle φ . We can write the scalar product appearing in the integrand as:

$$\mathbf{r}^T \cdot \mathbf{O} \cdot (\mathbf{A} \cdot \mathbf{r} + \mathbf{a}) = \text{Tr}[\mathbf{O} \cdot \mathbf{A}(\mathbf{r} + \mathbf{a}) \cdot \mathbf{r}^T] \quad (58)$$

and use the Schwarz inequality for functions on $[0, 2\pi]$ to arrive at the upper bound:

$$\begin{aligned} \frac{1}{\sqrt{2}} \int_0^{2\pi} \frac{d\varphi}{2\pi} \sqrt{(1 + \mathbf{a}^T \cdot \mathbf{r})^2 + (1 + \mathbf{a}^T \cdot \mathbf{r}) \text{Tr}[\mathbf{O} \cdot \mathbf{A}(\mathbf{r} + \mathbf{a}) \cdot \mathbf{r}^T]} \\ \leq \frac{1}{\sqrt{2}} \sqrt{(1 + \mathbf{a}^T \cdot \mathbf{r})^2 + (1 + \mathbf{a}^T \cdot \mathbf{r}) \text{Tr}(\mathbf{O} \cdot \mathbf{A} \cdot \mathbf{V})} \end{aligned} \quad (59)$$

where the matrix \mathbf{V} is given by

$$\mathbf{V} = \int_0^{2\pi} \frac{d\varphi}{2\pi} (\mathbf{r} + \mathbf{a}) \cdot \mathbf{r}^T = \text{diag}(\frac{1}{2} \sin^2 \theta, \frac{1}{2} \sin^2 \theta, x \cos \theta + \cos^2 \theta) \quad (60)$$

Let us now look for a rotation \mathbf{O} that will maximize the right hand side of the inequality (59). As \mathbf{O} enters this expression only through the trace $\text{Tr}(\mathbf{O} \cdot \mathbf{A} \cdot \mathbf{V})$ and furthermore both \mathbf{A} and \mathbf{V} are invariant with respect to rotations about the z axis, it is sufficient to consider rotations of the form $\mathbf{O} = \mathbf{R}_z(\alpha) \cdot \mathbf{R}_x(\beta)$. The trace written explicitly takes the form:

$$\text{Tr}(\mathbf{O} \cdot \mathbf{A} \cdot \mathbf{V}) = \frac{1}{2} \sqrt{1 - x^2} \sin^2 \theta \cos \alpha (1 + \cos \beta) + (\cos^2 \theta + x \cos \theta) \cos \beta \quad (61)$$

It is seen that this expression reaches the maximum for $\alpha = 0$ irrespectively of the values of any other parameters. Setting $\cos \alpha = 1$ and introducing a new integration variable $t = \cos \theta$ for the polar angle θ we can write the expression for the Bures-Uhlmann fidelity in the form of an integral:

$$\frac{1}{4} \int_{-1}^1 dt \sqrt{2(1 + xt)^2 + (1 + xt)[\sqrt{1 - x^2}(1 - t^2)(1 + \cos \beta) + 2(t^2 + xt) \cos \beta]} \quad (62)$$

It can be verified by numerical means that for any value of $x \in [0, 1]$ this integral reaches its maximum value for $\beta = 0$, corresponding to $\hat{U} = \hat{\mathbb{1}}$. When $\beta = 0$ the above integral can be evaluated analytically, with the final result given in (48).

Bibliography

1. A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, 1995).
2. A. K. Ekert, B. Huttner, G. M. Palma, and A. Peres, *Eavesdropping on quantum-cryptographical systems*, Phys. Rev. A **50**, 1047 (1994).
3. C. A. Fuchs and A. Peres, *Quantum-state disturbance versus information gain: Uncertainty relations for quantum information*, Phys. Rev. A **53**, 2038.
4. H. N. Barnum, *Quantum information theory*, Ph.D. thesis (University of New Mexico, Albuquerque, 1998).
5. K. Banaszek, *Fidelity balance in quantum operations*, Phys. Rev. Lett. **86**, 1366 (2001).
6. C. A. Fuchs and K. Jacobs, *Information-tradeoff relations for finite-strength quantum measurements*, Phys. Rev. A **63**, 062305 (2001).
7. K. Banaszek and I. Devetak, *Fidelity trade-off for finite ensembles of identically prepared qubits*, Phys. Rev. A **64**, 052307 (2001).
8. H. Barnum, *Information-disturbance tradeoff in quantum measurement on the uniform ensemble*, e-print quant-ph/0205155.
9. K. R. W. Jones, *Principles of quantum inference*, Ann. Phys. **207**, 140 (1991).
10. K. R. W. Jones, *Fundamental limits upon the measurement of state vectors*, Phys. Rev. A **50**, 3682 (1994).

11. V. Bužek, R. Derka, G. Adam, and P. L. Knight, *Reconstruction of quantum states of spin systems: From quantum Bayesian inference to quantum tomography*, Ann. Phys. **266**, 454 (1998).
12. M. A. Nielsen and C. M. Caves, *Reversible quantum operations and their application to teleportation*, Phys. Rev. A **55**, 2547 (1997).
13. T. Ando, *Concavity of certain maps on positive definite matrices and applications to Hadamard products*, Linear Algebra and its Applications **26**, 203 (1979).
14. C. Davies, *All convex invariant functions of hermitian matrices*, Archiv der Mathematik **8**, 276 (1957).
15. S. Friedland, *Convex spectral functions*, Linear Multilinear Algebra **9**, 299 (1981).
16. F. Jarre, *Convex analysis on symmetric matrices*, in *Handbook of Semidefinite Programming*, ed. by H. Wolkowicz, R. Saigal and L. Vandenberghe (Kluwer, Boston, 2000).
17. J. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization: Theory and Examples* (Springer, New York, 2000).
18. T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
19. K. Banaszek, *Optimal quantum teleportation with an arbitrary pure state*, Phys. Rev. A **62**, 024301 (2000).
20. R. Horst, P. M. Pardalos, and N. V. Thoai, *Introduction to Global Optimization* (Kluwer, Dordrecht, 1995).
21. K. R. W. Jones, *Riemann-Liouville fractional integration and reduced distributions on hyperspheres*, J. Phys. A **24**, 1237 (1991).
22. S. Massar and S. Popescu, *Optimal extraction of information from finite quantum ensembles*, Phys. Rev. Lett. **74**, 1259 (1995).
23. R. Derka, V. Bužek, and A. K. Ekert, *Universal algorithm for optimal estimation of quantum states from finite ensembles via realizable generalized measurement*, Phys. Rev. Lett. **80**, 1571 (1998).
24. G. Vidal, J. I. Latorre, P. Pascual, and R. Tarrach, *Optimal minimal measurements of mixed states*, Phys. Rev. A **60**, 126 (1999).
25. D. Bures, *An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w^* -algebras*, Trans. Amer. Math. Soc. **135**, 199 (1969).
26. A. Uhlmann, *The "transition probability" in the state space of a $*$ -algebra*, Rep. Math. Phys. **9**, 273 (1976).
27. M. Hübner, *Explicit computation of the Bures distance for density matrices*, Phys. Lett. A **163**, 239 (1992).
28. R. Jozsa, *Fidelity for mixed quantum states*, J. Mod. Opt. **41**, 2315 (1994).